Design Methodologies, Models and Tools for Very-Large-Scale Integration of NEM Relay-Based Circuits

Tian Qin, Sunil Rana and Dinesh Pamunuwa
Dept. of Electronic & Electrical Engineering, University of Bristol, Bristol, UK BS8 1UB
eextq@bristol.ac.uk

ABSTRACT
Integrated circuits based on nano-electromechanical (NEM) relays are a promising alternative to conventional CMOS technology in ultra-low energy applications due to their (near) zero stand-by energy consumption. Here we describe the details of an overarching design framework for NEM relays, including automated synthesis from design entry in RTL to layout, based on commercially available EDA tools and engines. Critical differences between relays and FETs manifest in fundamentally different timing characteristics, which significantly affect static timing analysis and the requisite timing models. The adaptation of existing EDA methods, models, tools and platforms for logic and physical synthesis to account for these differences are described, providing insight into large-scale design of NEM relay-based digital processors. A historically well-known processor, the Intel 4004, and a modern MIPS32 compatible processor are synthesized based on a NEM relay-based standard cell library to demonstrate the customized synthesis methodology. An energy study is carried out using the proposed design framework on benchmark circuits implemented in existing CMOS nodes and NEM node, to better understand the energy saving potential of NEM technology.

Keywords
NEM relay, synthesis flow, timing analysis, standard cell characterization, design framework

1. INTRODUCTION
Reduction of the minimum feature size has been at the heart of the unparalleled success of CMOS technology, historically providing simultaneous improvements in the propagation delay, energy consumption and footprint of a binary switching transfer – the canonical digital computing operation. In nanometer technologies, subthreshold conduction has become a significant issue, and carrier statistics dictate a lower limit on the subthreshold swing of 60 mV/decade at room temperature [1]. Thus the reduction in dynamic energy achievable by shrinking the rail voltage is offset by increased leakage energy [2]. NEM relays, due to their zero off-state current and an abrupt on/off transient, hold out the promise of an energy efficiency unattainable by MOSFETs [3]. There has been considerable interest in NEM relay based digital circuits recently and several device primitives have been reported [4-7]. Functional digital gates and circuits based on in-plane as well as out-of-plane relays have also been reported [8,9]. While stiction in the contact and overall reliability are major issues, significant progress has been made [10].

As NEM relay technology has matured, attention to hierarchical modelling [9,11,12] and synthesis frameworks [13] has increased. In this paper we report a fully automated top-down design framework based on commercially available EDA tools that supports logic synthesis, placement and routing from RTL and schematic entry. This framework is underpinned by a hierarchical set of simulation models comprising physical models derived from model-order reduction of finite element models that can be incorporated in spice simulations, and behavioral models that enable event-driven digital simulations. The top-level design is captured in terms of elements available in a standard cell library, where each cell has been characterized according to a bespoke methodology devised to accommodate the unique timing characteristics of NEM relays. This characterization methodology allows accurate static timing analysis –which is at the heart of any timing constrained synthesis– even though NEM relays exhibit hard to predict delay variation depending on the possibility of actuating the relay at some random point in the free oscillation of the beam after release [14].

The development of the design framework, including the details of NEM device modelling flow, gate-level timing characterization and automated synthesis flow, is presented in Section 2. Synthesis of the Intel 4004 processor [19] and a 32-bit MIPS processor [15] using the developed framework, as well as key metrics such as critical path latency, gate and device count and area is described in Section 3. A case study on energy consumption of NEM technology is carried out in the same section. The conclusions are presented in Section 4.

2. DESIGN FRAMEWORK

2.1 Overview
The NEM relay-based VLSI circuit design flow that has been developed is a standard-cell based semi-custom top-down design flow. It utilizes commercially available EDA tools and engines used for CMOS IC design. All device models and technology files used in each step of the design flow need to be customized for the target NEM technology. Design capture is based on a NEM relay-based standard-cell library. The standard cells are functionally verified and characterized through analog simulation using measurement-qualified device models. The physical layout of the standard cells is realized using a full-custom approach in the Cadence Virtuoso design environment. Cadence RTL Compiler is used for logic synthesis and Cadence Encounter Digital Implementation (EDI) Platform is used for automated placement and routing. A flowchart illustrating the entire flow as well as the customization that takes place is shown in Fig. 1. Overall, the most critical customization for the target technology comprises modification on the following three files:

1. a Cadence Design Framework II (DFII) technology file that provides technology information to Virtuoso including the layer definitions, physical and electrical rules, and allowed polygon and separation dimensions that define the design rules for the process;
2. a liberty file used by the synthesis tool that contains critical timing and energy information for each cell;
3. a library exchange format (LEF) file that describes the physical geometry and pin information for each cell, which is necessary for automatic placement & routing.

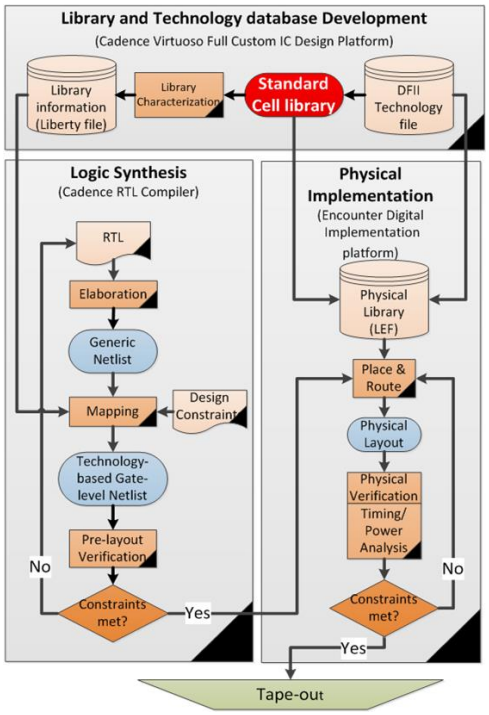

Figure 1: Semi-custom and automated design flow showing customization for target technology

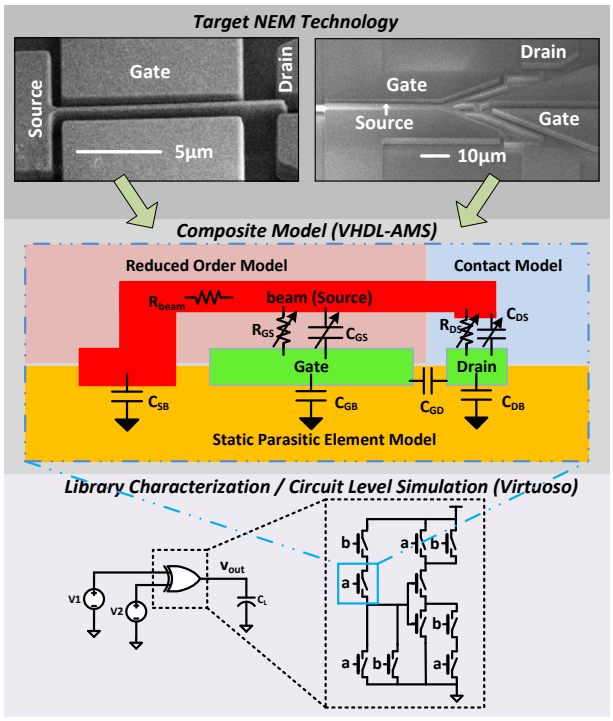

Figure 2: Top: two NEM technologies (Left: Fabricated relay from [6]; Right: our prototype relay); Middle: Relay model comprising reduced-order model for electro-mechanical beam dynamics, contact model for surface mechanics and electrical behavior and static parasitic element model; Bottom: Circuit Simulation based on the device model (where an XOR2 gate is shown as example);

Of these, the textual DFII technology file is the primary definition of process-related details. It is compiled into a technology library, which when attached to the cell library, allows custom layout design with DRC and extraction. The liberty file is generated from characterization of the standard cells based on the accurate NEM relay device model (see section 2.2.3) and underpins logic synthesis.

It should be noted that this design methodology is technology agnostic, and same procedure can be followed for any NEM technology, or indeed any disruptive technology, where the physical details of the process can be captured in terms of a set of design rules for the material layers contained in the process, and the device behavior described by a mixed-mode or equivalent circuit model. It is also worth noting that the design framework is capable of carrying out CMOS-NEM heterogeneous design because in most NEM processes, the relays are either fabricated on top of CMOS on the same wafer, or on another wafer followed by back-end-of-line (BEOL) integration with the CMOS wafer. This effectively means that NEM and CMOS technology occupy different physical layers allowing layer-specific separation of design rules associated with each technology, in the EDA tool. Hence the NEM and CMOS tech files can be integrated into a single file without causing conflict. With such capability, the proposed design framework can potentially be used for increasing design productivity for any heterogeneously integrated CMOS-NEM design (e.g. [20]), since everything on the chip is designed in the same environment.

2.2 Device Modeling, Simulation and Characterization

2.2.1 Physical Modeling and Analog Simulation

A technology agnostic physical modelling flow is used to generate the device model in the framework, and a specific NEM relay technology has been used as an example. To accurately describe the behavior of a single relay in all regimes of operation from closed to open including transient behavior, a composite model comprising three separate components – an electromechanical model describing the beam dynamics, a contact model describing surface contact mechanics and electrical behavior before and after source to drain (stationery electrode) physical contact, and a parasitic element model – is used.

To model the gate-source electromechanical behavior, finite-element analysis (FEA) is carried out using a solid model of the relay. A reduced-order model (ROM) is then developed based on the FEA results through the method of model order reduction. The ROM includes polynomial fitting functions, and specifies the transient electromechanical behavior of the relay cantilever far more accurately than is possible with the ubiquitous parallel-plate capacitor model, with the non-linear mechanical bending of the beam accurately captured [16].

The drain-source contact model accounts for the surface interactions between the source tip and the drain – van der Waals dispersion forces and repulsive force of the electron clouds – using the Lennard-Jones potential function [14]. It also accounts for the drain-source contact resistance and capacitance. The tunneling current flowing between the source tip and the drain at atomic level separations is modeled using a transconductance function [17]. Once a physical source-drain contact is established, the contact resistance is dictated by the effective contact area and the interfacing materials. The contact capacitance is modelled using a parallel plate approximation that contains a correction function to

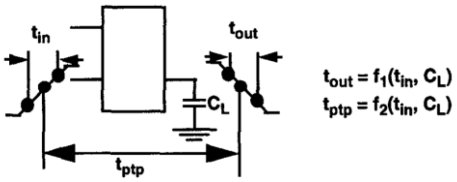

Figure 3: Traditional Characterization of CMOS gate [18]

reduce the capacitance to zero as the resistive current becomes dominant. This capacitance acts to reduce the relay pull-in voltage, similar to the drain-induced barrier lowering (DIBL) effect in MOSFETs.

The final component of the model accounts for the parasitic capacitances and resistances present in the device. These are extracted using a commercial extraction engine (Raphael from Synopsys), and comprise static values. The ROM and contact model have been implemented in VHDL-AMS, while the parasitic components have been incorporated as circuit elements. The full model is shown in Fig. 2.

The model also incorporates variation in the critical physical parameters of air gap, beam and hinge dimensions, which translate to variations in pull-in and pull-out voltages and relay closing time. The amount of parametric variation can be chosen at the time of model instantiation, allowing corner analysis to be performed. The corners are defined by minimum, nominal and maximum values for physical dimensions and represent process variation.

This model can be used for circuit level simulation and is compatible with the mixed-mode simulation environment available in all major EDA tools. In our circuit design experiments we have used the Spectre simulator in the Cadence Analog Design Environment.

2.2.2 Gate-level Modeling for Digital Simulation and Top-down Synthesis

A standard cell library is developed based on a prototype three-terminal architecture (see Fig. 2, top right). A scaled version of this relay has a nominal delay of ~50 ns and a footprint of ~5 μm^2. The cell-library has inverters, buffers, tristate buffers, NAND, NOR, XOR and XNOR gates, full-adders, D latches, D flip-flops and multiplexers as primitive cells. All these cells have been designed in a complementary style with pull-up/-down networks to accommodate the three-terminal devices [14].

The full analog behavioral model described above provides the benchmark for accuracy, and is used in full-custom circuit design and verification. It is however too time consuming to run an analog simulation for full chip-level verification of large designs. Further, automated synthesis requires abstract gate-level models with timing and energy consumption specified against input signal characteristics and output load. This allows the propagation delay associated with a network of gates to be abstracted by timing arcs, i.e. the delay associated with signal flow from a given input pin to a given output pin.

In static timing analysis, the delay along a path is obtained through adding up the delays of the timing arcs forming the path. In the characterization of a CMOS standard cell library, the gate delay is traditionally customized as a function of input slew rate and output capacitive load as shown in Figure 3 (from [18]). Characterization of NEM libraries requires a more complex approach, as the delay associated with driving a load comprises the mechanical transition delay (time taken by the relay beam tip to traverse the contact gap) and the electrical *RC* delay of the circuit.

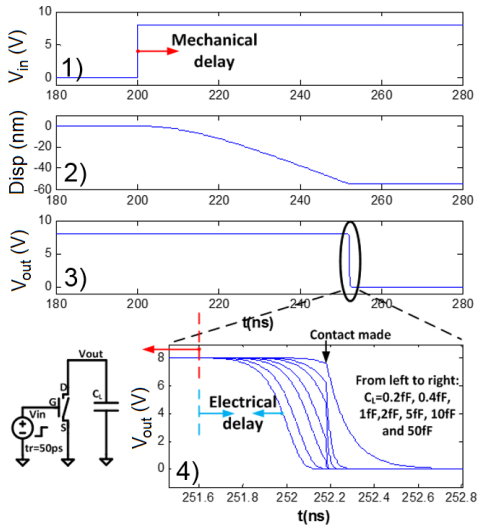

Figure 4: 1) Actuation signal, 2) relay beam tip displacement, 3) output node voltage and 4) high resolution trace of voltage signal transition at the output node as the relay closes.

For all recently reported NEM relays [4-7,9] the mechanical delay is the dominant part of the total delay, typically more than 2 orders of magnitude larger than the *RC* delay with an on-resistance of the order of 10-20 kΩ and load capacitances in the tens of fF range. Consequently, most works (e.g. [13, 21]) on NEM circuit design and synthesis use a constant mechanical delay value as the overall switching delay.

This assumption breaks down for two reasons. Firstly, the mechanical delay of a NEM relay (especially those with a high Q) is not constant. It varies depending on how soon the relay is re-actuated, because a de-actuated relay beam oscillates until the stored potential energy of the beam is dissipated through damping. Thus, the switching-on time is dependent on the position and velocity of the beam at the time of actuation. Secondly, as NEM relay technology matures and device size is scaled down, its mechanical delay falls into sub-nanosecond range and electrical delay will eventually become comparable [8], and this is especially true when interconnection parasitics are taken into account. Even for existing relay designs, when a gate has a large fan-out, the input capacitance of the next stage gates and the metal interconnection make the RC delay non-trivial.

Figure 4 illustrates the complex nature of the transition delay of a relay. When the actuation signal (V_{in}) is applied across the gate-source capacitor (trace 1) the relay beam (source) starts to move towards the drain contact (trace 2). As the relay closes (i.e. drain-to-source contact is established) the capacitive load (pre-charged to V_{dd} – 8V) discharges (traces 3 and 4). The multiple signals in the high resolution trace 4 (corresponding to capacitive loads of 0.2fF, 0.4fF, 1fF, 2fF, 5fF, 10fF and 50fF) reveal the true nature of the electrical transition at the output node. Due to the presence of tunnelling at sub-nanometre separation between the beam tip and the drain a current starts to flow before a physical drain-source contact is established. Due to the non-linear nature of the variation of tunneling current with separation, the discharging up to this point is not characterized by typical *RC* behavior. Once an ohmic contact is established between the source and drain, typical RC behavior can indeed be seen.

Based on the above behavior, the total transition delay is separated into 1) an intrinsic delay where the tunneling current is negligible

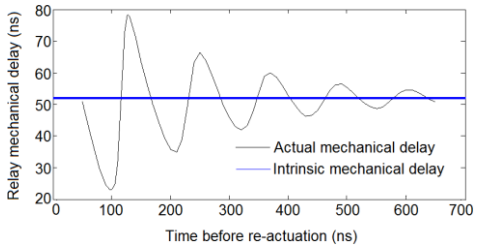

Figure 5: Variation of mechanical delay with increasing time intervals between de- and re-actuation.

and 2) an extrinsic (electrical) delay where the load starts to experience charging/discharging. The intrinsic delay is defined as the time required for the relay beam tip, when actuated from rest (using an ideal step-signal), to reduce the source-drain separation to 1 nm; its value is constant. The electrical delay, on the other hand, is determined by the load capacitance and may range from hundreds of picoseconds to a few nanoseconds (depending upon the load). This separation of the mechanical and electrical delays enables accurate static timing analysis as the variation in both (see below for variation in mechanical delay), which are governed by different mechanisms, can be addressed separately.

The definition of the intrinsic delay above corresponds to relay actuation from rest using an ideal step signal. If, however, the beam is oscillating at the instant of actuation, the mechanical delay can be much greater or smaller that the intrinsic delay. For example, based on device simulations, when the applied actuation voltage is 8V, the nominal mechanical delay of a 3-terminal relay is 51.6 ns, and the best case and worst case delays are 22.8 ns and 77.8 ns respectively (Figure 5) (see [14]). The worst-case mechanical delay is approximately 4× the best-case. As expected, the mechanical delay gets progressively closer to the intrinsic delay as the actuation frequency is reduced. When the time difference between de-actuation and re-actuation is greater than 600ns, the variation is less than 10% of the intrinsic delay. This is because the free oscillation gradually dies down due to energy loss through damping (the 3-terminal relay considered for the analysis has a Q-factor of 55). This phenomenon has a profound effect on calculation of both the worst-case latency in individual gates and the critical path delay.

With 3-terminal NEM relay technology, relays in a series path in the pull-up (down) network turn on sequentially, as a potential difference between the beam and the control electrode (gate terminal) for a given relay is only established when the relay above (below) it turns on. In the worst-case though, for a single gate, the propagation delay is **not** $N*t_{pd\_wc}$ where N is the number of gates in series and t_{pd_wc} the worst-case propagation delay of a single device. This is because even if the arrival time of the gating signal in a new cycle corresponds to the worst-case, due to the above effect, only the relay closest to V_{dd} (ground) in the pull-up (down) network will ever see the worst-case interval between de- and re-actuation. Furthermore, for the same reason, the effect of free oscillations on the gate-level propagation delay decreases with the number of gates in a series path of combinational gates.

Figure 6 shows the variation of the path delay of an M-stage inverter chain where the worst-case and best-case delays have been obtained by carefully engineering the actuation pulses so as to hit the worst-/best-case timing points. As an inverter has only a single relay in the pull-up and -down networks, the worst-case mechanical delay of an inverter is identical to that of a relay. The simulations show that estimating the path delay by $M*t_{pd\_wc}$ is overly pessimistic while $M*t_{pd\_bc}$ is too optimistic; the longer the chain,

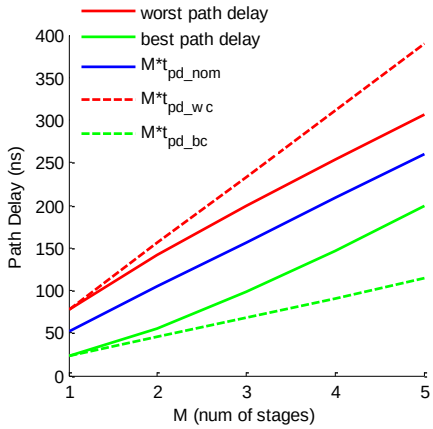

Figure 6: Variation of total delay of an inverter chain with increasing number of stages.

the closer the overall propagation delay to $M*t_{pd\_nom}$. This is clearly because the longer the chain, the more time elapses when the actuation signal propagates to later stages. Thus the oscillation amplitude of devices in later stages reduces through damping by the time they are re-actuated. When the number of stages is high enough, the oscillation-based delay effect is only seen in the first few stages (or relays), and the later stages tend to have the intrinsic gate delay. Thus the traditional way of calculating the critical and contamination delay as the sum of the worst-case and best-case propagation delays can result in significant errors, and a more accurate measure is obtained by using the intrinsic stage delay for path delay calculation in static timing analysis, modified by an empirical correction factor to account for oscillation-based timing variation in the first few stages.

2.2.3 Digital Timing Model

The total propagation delay, D, of a NEM logic gate is defined as the sum of the mechanical delay, D_M, and electrical delay, D_E.

$$D = \alpha D_M + D_E \qquad (1)$$

Here α is an empirically estimated correction factor to account for the effect of oscillation-based delay variation, which can take one of two values depending on whether the propagation or contamination delay is being estimated. D_M is the nominal mechanical delay, i.e. the delay when the beam is actuated from rest with a finite slope. It is affected by the input signal slew rate and is further divided into 1) an unchanging intrinsic delay (D_I), corresponding to an ideal step actuation, and 2) a slew rate dependent component (D_S):

$$D_M = D_I + D_S \qquad (2)$$

D_S is defined as the product of a slope sensitivity factor, S_S, and the transition delay calculated at the output pin of the previous stage, D_{T_prev}:

$$D_S = S_S \times D_{T\_prev} \qquad (3)$$

D_I, which is constant for a given voltage step input, is determined using a full analog simulation of the logic gate while D_M is measured against different values of D_{T_prev}. D_S and hence S_S is obtained through linear fitting based on (2) and (3). D_I is measured as the time interval between the 50% point of the input signal (v_{in}) and the instant when the beam tip-to-drain tunnelling becomes significant (corresponding to a tip-to-drain separation of 1 nm for the targeted technology). The characterization of S_S for an inverter is shown in Figure 7.

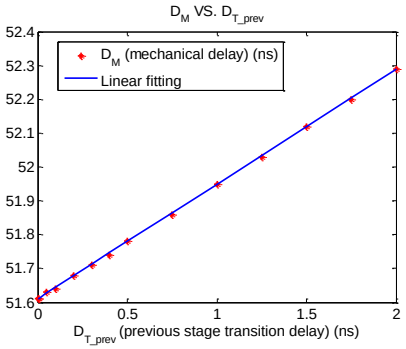

Figure 7: Characterization of the slope factor Ss

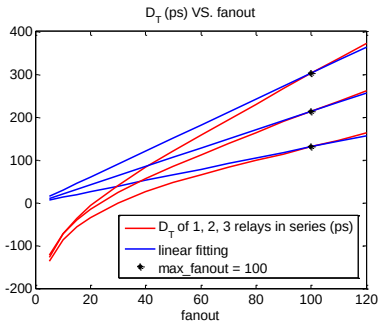

Figure 8: Characterization of R_{drv}

The electrical delay D_E is determined by the load being driven by the gate under characterization. To characterize the electrical delay, varying loads (from 1 to 20 NEM stages as well as static capacitors from 0.2 fF to 50 fF) are driven, and a table of load vs. electrical transition delay is generated through high resolution simulations. The electrical delay is the sum of the *RC* delay associated with the load, D_T, and connect delay, D_C:

$$D_E = D_T + D_C \qquad (4)$$

For negligible connect delay $D_E = D_T$, where D_T is determined by the equivalent resistance of the driver, R_{drv}, and load capacitance, C_{load}. Thus, the electrical delay is modeled as follows where R_{drv} is the key quantity that needs to be characterized.

$$D_E = D_T = R_{drv}\, C_{load} = R_{drv}\, (C_{wire} + C_{pins}) \qquad (5)$$

For consistency, the D_T values are measured from the instant when the source beam tip to drain separation is 1 nm (i.e. when the tunneling current starts flowing) to the instant when the output signal reaches 50% of its maximum. Figure 8 illustrates the characterization of R_{drv} for the pull-down network of an inverter, NAND2 and NAND3. The deviation seen when the fan-out is low is caused by measuring D_T at 1 nm separation. Under a very low fan-out, the tunneling current that flows before the separation reaches 1 nm, charges up the very small load capacitance. Hence the calculated electrical delay is pessimistic. This variation at low fan-out is not an issue, as the total load is equivalent to a fan-out of 40-100 with the interconnection load taken into consideration, in a typical NEM technology.

The proposed definition of the total switching delay in terms of the electrical and mechanical delays makes the addition of correction factors, to generate the best and worst case values from the nominal ones, significantly simpler. A look-up table based approach by contrast requires a much higher effort as all the gates in the library would need to be individually characterized through high resolution simulations at multiple voltage and process corners. Besides, the proposed definition of the total switching delay in terms of the electrical and mechanical delays makes the addition of correction factors, to generate the best and worst case values from the nominal ones, significantly simpler.

The final enabling step in developing a synthesis capability is the generation of a technology library for synthesis through incorporating the characterized delay models into the liberty files.

2.2.4 Digital Power Model

Traditionally, the power characterization of a logic gate is categorized into: (1) static power, which is mainly caused by subthreshold leakage in CMOS when the gate is inactive and (2) dynamic power, which comprises short-circuit power and dynamic switching power related to charging/discharging of the load capacitance. Here the static power is modeled as zero as NEM relays have zero leakage.

NEM relays experience hysteresis, with the pull-in voltage (V_{pi}) always greater than the pull-out voltage (V_{po}) by an amount V_h, ie. $V_{pi} = V_{po} + V_h$. In a complementary style implementation with fault-free relays (i.e. stiction does not affect the mechanical pull-out), as long as $V_{DD} < 2V_{pi} - V_h$, pull-out of a relay in either the pull-up or pull-down network is guaranteed to take place before pull-in occurs in its counterpart in the other network. Thus the pull-up and pull-down networks are never on at the same time for fault-free relays, and the short-circuit power is also modelled as zero.

The switching power is modelled as $0.5 C_{load} Vdd^2 f_p$, where C_{load} is the sum of the downstream net and gate capacitances and f_p is the activity factor modified switching frequency.

3. TIMING-DRIVEN SYNTHESIS

3.1 Demonstration of the Design framework

Synthesis from RTL design entry to layout has been carried out on two processors using the developed design framework, the first example being the first commercially-available single-chip CPU, the Intel i4004 processor. Both the datapath and controller of the i4004 was realized using timing-constrained synthesis using the developed fully-automated flow, using a Verilog RTL description of the processor for design entry [19]. Details of the synthesized i4004 along with details of the original implementation are given in Table 1. The results indicate the example NEM technology achieves similar performance to the 1st-generation PMOS technology with ~3x savings on area. Since the NEM i4004 datapath is a synthesis effort, the performance of NEM i4004 can be greatly improved with a customized datapath.

Table 1. Synthesized i4004 processor

Technology	1st-gen Self-aligned p-channel MOSFET	Target NEM
Area	~12 mm^2 (core + pad)	~4.5 mm^2(core)
Dimension	3.0 mm x 4.0 mm	2.1mm x 2.2mm
Max. Clock Frequency	750kHz	645kHz
Min instruction cycle	10.7us	12.4us
Number of Devices	~2,300 pMOSFETs	~2000 Logic Gates
Supply Voltage	-15V (or -12V to +5V)	8V

Another example to demonstrate the synthesis capability of the framework is a 5-stage pipelined 32-bit MIPS processor [15], implemented with full data forwarding and hazard detection. The synthesis result is shown in Table 2. It can be seen that the NEM implementation of the MIPS32 processor uses more gates than its 0.35µm CMOS counterpart, and this is because the example NEM standard cell library is less diverse than its CMOS counterpart.

Table 2. Synthesized MIPS32

Technology	0.35µm CMOS	Target NEM
Area	~3mm^2(core)	~ 33mm^2(core)
Max Clock Frequency	100MHz	0.5 MHz
Number of instances	~26K gates	~64Kgates
Power	0.91W	0.06W
Supply Voltage	3.3V	8V

3.2 Timing Verification

The critical issue to be determined is the validity of using an empirically determined static delay for gates to account for delay variation caused by free oscillation. For this purpose, static timing analysis carried out on the critical paths identified from the netlist of the synthesized processors (e.g. Figure 9) has been compared with the delays obtained through full analog simulations.

As discussed earlier, generally the shorter a critical path, the larger the mechanical delay variation when compared with the overall path delay. Shown in Figure 10 are the path delays along a 7 mechanical delay data path extracted from the synthesized netlist of the i4004 processor. The path delays are extracted from high-resolution analog simulations for 50 switching frequencies linearly spaced between the low and high frequencies corresponding to the path delays associated with each stage experiencing the worst- and best-case delay respectively. The x-axis shows the delay normalized to the path delay associated with each stage experiencing the *intrinsic* delay. Each simulation runs for many cycles, and a stable state is reached where the delay stabilizes, if the switching frequency is within the bounds that the path can operate in. For each run, the best-case, worst-case and stable steady-state (i.e. where the delay converges to some value after several cycles) delays that occur over the entire run have been extracted. These delay values are shown in separate histograms. As can be seen, for this 7 mechanical stage path, the worst-case never exceeds 110% of the nominal delay, while the best-case does not drop below 85%. This verifies our previous assertion that estimating the critical path delay and the contamination delay by using the worst- and best-case stage delays would result in massive over- and under-estimations respectively, which is around 56% and 51% for best- and worst-case respectively. By contrast, using an empirical correction factor to account for free-oscillation-based delay variation provides a high degree of accuracy.

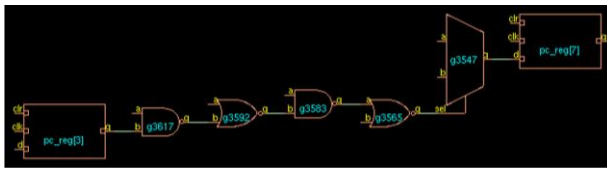

Figure 9: A critical path in the synthesized i4004 based on the experimental NEM technology

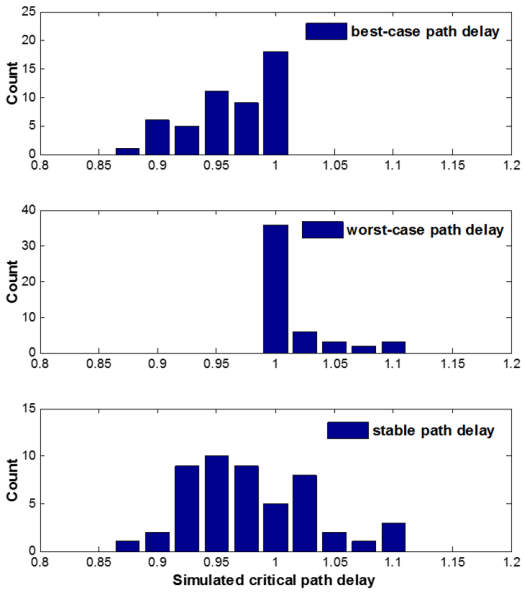

Figure 10: Simulated Critical Path Delay (Normalized against the result calculated from static timing analysis)

Shown in Table 3 are examples of the worst-case delays extracted from analog simulations for longer data paths. When the number of mechanical delays along the critical path goes up to 10, the worst-case delay never exceeds 10% greater than the nominal path delay. Thus for this specific technology, as long as the critical path on chip has more than 10 stages of mechanical delay, it is safe to use $110\%*M*t_{pd\_nom}$ for critical-delay estimation, where M is the number of mechanical delay along the path and $M \geq 10$.

Table 3. Extracted Worst Case Delay of Logic Paths

Logic Path	No. of nominal mechanical delays in path	Worst case delay normalized to nominal delay
A	10	1.030
B	10	1.014
C	9	1.089
D	9	1.078
E	8	1.027

It is worth noting though, although the proposed characterization methodology and synthesis flow is technology agnostic, the empirical correction factor will vary with a different NEM technology, as the Q factor of the device will in general be different.

3.3 Case study on energy and performance of NEM Relay-Enabled Logic

To understand the opportunities for deployment of the NEM relay-based technology, comparison with state-of-the-art CMOS is necessary. To this end, we study the post synthesis results from the i4004 implemented in the modelled NEM relay technology and two other commercially available CMOS processes at 65 nm and 0.35 µm technology nodes respectively.

The synthesis result is highly affected by user defined constraints related to timing and area, as well as the diversity available in the standard cell library (i.e., types of combinational and sequencing elements available). To ensure the comparison is a fair one, for each technology, synthesis is carried out using only the same subset from the corresponding cell library, under loosely defined timing

Technology	65 nm CMOS		0.35 um CMOS		5 um NEM with 60 nm air gap		4T Hypothesis NEM (body-biased)	
Column	①	②	③	④	⑤	⑥	⑦	⑧
Operating Voltage (V)	1.3	1.3	3.3	3.3	8	8	±1.65	±0.65
Area (um^2)	3982.7	3982.7	137443	137443	1579770	1654144	1654144	1654144
Critical Delay (ns)	1.75	1.75	6.12	6.12	1751.9	1754.2	1754.2	1754.2
Clock Frequency (Hz)	20M	500k	20M	500k	500K	500K	500K	500K
Wire load Model	65 nm	65 nm	0.35um	0.35um	N/A	0.35um	0.35um	0.35um
Energy per Cycle (nJ)	**0.0058**	**0.1755**	**0.218**	**0.218**	**0.133**	**2.298**	**0.391**	**0.06**
Total Power (mW)	0.1163	0.0877	4.36	0.109	0.0665	1.149	0.196	0.03
Dynamic Power (mW)	0.0293	0.0007	4.36	0.109	0.0665	1.149	0.196	0.03
Leakage Power (mW)	0.087	0.087	1E-06	1E-06	0	0	0	0

Table 4. Post-synthesis results for i4004 in NEM and CMOS technology

constraints and no power / area constraints. This ensures that the synthesis result achieves the lowest power consumption at the minimum area, and the adverse effects caused by the example NEM cell library being less diverse is eliminated. Since the NEM circuit has to operate under a much lower clock frequency than CMOS due to the inherently large mechanical delay, comparison based on power dissipation alone doesn't yield very meaningful result. For a fair comparison on the energy/power consumption, we consider the average energy consumption per clock cycle, which is the closest metric to average energy consumption per operation.

Table 4 shows the post-synthesis results of the i4004. Column 1 through 4 are results from implementations in the 65 nm and 0.35 µm CMOS technology nodes when the processor is operating at clock frequencies of 20 MHz and 500 kHz respectively. Columns 5 and 6 are results from the considered NEM technology operating at a 500 kHz clock frequency. Columns 7 and 8 are results from a hypothetical 4-Terminal NEM technology operating at a 500 kHz clock frequency when the body-bias technique [14] is applied. The assumption is the 4-terminal device has an identical footprint and material properties as the 3-terminal device used in this study.

When considering power/energy consumption of NEM relay-based circuits, prior works mostly ignore the effect of interconnects, which led to overly optimistic predictions on the energy benefit of NEM relay-based logic. In this work, interconnection parasitics have been taken into account for more realistic energy prediction by including a wire-load model in the estimation. A wire load model is a statistical model based on previous fabricated chips that estimates the interconnection parasitics (*R, C* based on length and area) based on gate fan-out. A 0.35 µm 3-metal wire load model is used for this study, under the assumption of interconnection of NEM relays based on BEOL integration with this technology.

The following observations can be drawn from the results in Table 4: Firstly, when both NEM and CMOS circuits are operating in conventional mode with no energy saving techniques applied and at their natural operating frequency (20MHz and 500KHz respectively, see column 1, 3 and 5), the current NEM relay we are focusing on (~5 μm^2 footprint, ~50 ns mechanical latency and ~7V pull-in voltage) doesn't have an advantage in energy consumption over deep sub-micron CMOS (see column 1 and 5), though an energy saving of around 40% is evident when compared with the 0.35 µm CMOS technology (see column 3 and 5). However, when both the CMOS and NEM implementations operate at very low frequencies (column 2, 4 and 5), for 65 nm CMOS technology, the leakage energy per cycle increases drastically and starts dominating the overall energy consumption. The older technology CMOS node (0.35um), on the other hand, doesn't see an obvious increase in energy consumption due to its high threshold voltage and low leakage. Thus NEM relay technology shows a clear advantage over deep sub-micron CMOS technology in low frequency operation (see column 2 and 5), due to its inherent zero-leakage. This conclusion is consistent with results in previous studies. However these results are based on estimations that ignore the effect of interconnects in the NEM circuits (column 5). It should also be noticed that, in this experiment, to ensure the synthesized netlists in the three technologies are as similar to each other as possible, common low-power techniques such as power gating and clock gating are not used. While such techniques have no appreciable effect on NEM and older CMOS technologies operating at their most natural operating frequencies, their effect on deep sub-micron technologies (such as the 65 nm node) operating at low frequencies is profound. Deep submicron circuits invariably use energy saving techniques at low frequencies and operate in subthreshold mode in ultra-low power applications.

When a wire load model is used in the estimations for NEM circuits it can be seen that the total energy consumption of NEM implementations rises drastically, and the dynamic energy dissipation caused by switching activity on the interconnection nets actually becomes the dominant part of the total energy consumption (see column 5 to 6). Although it is well known that a wire load model only provides a limited degree of accuracy and tends to be overly pessimistic, the interconnection effect on overall energy consumption is a critical aspect that cannot be overlooked. This result is easily explained, as a given circuit footprint in a NEM technology with a device size of the order of ~5 μm^2 is much larger than in even early CMOS technologies. Hence the area of the synthesized NEM system will be much larger than CMOS and the length of the interconnection will be longer.

When the body-biasing energy saving technique is applied (see [14]), NEM relay technology shows up as having very promising energy saving potential (see column 7 and 8). The example given in column 8 shows that an energy saving of around 97.5% percent can be achieved when the body-biasing technique is applied and the input driving voltage to each stage is reduced from the full rail-to-rail swing (0V~8V) to just above the hysteresis window (-0.65V ~+0.65V). With body-biasing applied, when all three implementations operate at their natural frequency (20 MHz for CMOS and 500 kHz for NEM), it can be seen that the hypothetical 4-terminal NEM technology has a 72.5% energy saving in comparison to 0.35 µm CMOS and is able to compete with deep-sub micron CMOS, albeit at the expense of reduced noise margin (see column 1, 3 and 8). For these estimations, the effect of interconnects is taken into account for all technologies.

4. Conclusions

Significant differences in the physical behavior of NEM relays when compared to MOSFETs present various challenges in their usage to build logic circuits. In particular, delay variation is caused by the actuation signal arriving at variable points of the free oscillation of the beam after de-activation in a previous cycle. A custom timing model however provides insight into this effect, and enables automated synthesis with near optimal timing, a critical requirement given the relatively high mechanical delay of relays. Thus accurate capture of the physical behavior of NEM relays through a hierarchical set of simulation models allows the powerful capability of existing EDA platforms to be utilized for their large-scale integration. An energy study reveals that for current generation NEM technology, the dynamic energy dissipated on interconnections is the dominating factor of the total energy. This originates from the inherently large device footprint of NEM relays and makes it hard for NEM to compete with commercial CMOS at its current state of technological readiness. However, with improved architectures, specifically development of 4-terminal relays to enable body-biasing, and continuing scaling-down of device size, future generations of NEM relay technology still promise significant savings over CMOS in the domain of ultra-low energy applications.

5. Acknowledgement

We gratefully acknowledge financial support from the European Commission under the 7th Framework Programme (FP7) for the NEMIAC project (Grant no. 288670) which has in part enabled this work.

6. REFERENCES

[1] J. D. Meindl and J. A. Davis, "The fundamental limit on binary switching energy for terascale integration (TSI)," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1515-1516, Oct. 2000.

[2] B. H. Calhoun, A. Wang and A. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *IEEE J. Solid-State Circuits*, vol. 40, pp. 1778-1786, Sep. 2005.

[3] A. M. Ionescu, L. De Michielis, N. Dagtekin, G. Salvatore, J. Cao, A. Rusu and S. Bartsch, "Ultra low power: Emerging devices and their benefits for integrated circuits," in Proc. *IEEE Int. Electron Devices Meeting*, Washington, 2011, pp. 16.1.1-16.1.4.

[4] R. Nathanael, V. Pott, H. Kam, J. Jeon and T. K. Liu, "4-terminal relay technology for complementary logic," in *Proc. IEEE Int. Electron Devices Meeting*, Baltimore, MD, 2009, pp. 9.4.1-9.4.4.

[5] W. S. Lee, S. Chong, R. Parsa, J. Provine, D. Lee, S. Mitra, H-SP Wong and R. T. Howe, "Dual sidewall lateral nanoelectromechanical relays with beam isolation," *16th Int. Solid-State Sensors, Actuators and Microsystems Conf.*, Beijing, 2011, pp. 2606-2609.

[6] S. Chong, K. Akarvardar, R. Parsa, J. Yoon, R. T. Howe, S. Mitra and H-SP. Wong, "Nanoelectromechanical (NEM) relays Integrated with CMOS SRAM for improved stability and low Leakage," *IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD) – Digest of Technical Papers*, San Jose, CA, 2009, pp. 478-484.

[7] D. Grogg, U. Drechsler, A. Knoll, Y. Pu, C. Hagleitner, and M. Despont, "Curved cantilever design for a robust and scalable microelectromechanical switch," in *Proc. 56th International Conference on Electron, Ion, and Photon Beam Technology and Nanofabrication (EIPBN)*, Waikoloa, Hawaii, 2012.

[8] C. L. Ayala, D. Grogg, A. Bazigos, M. F. Badia, U. T. Duerig, M. Despont and Christoph Hagleitner, "A 6.7 MHz nanoelectromechanical ring oscillator using curved cantilever switches coated with amorphous carbon," *44th European Solid State Device Research Conf.*, Venice Lido, 2014, pp. 66-69.

[9] M. Spencer, F. Chen, C. C. Wang, R. Nathanael, H. Fariborzi, A. Gupta, H. Kam, V. Pott, J. Jaeseok, T. K. Liu, D. Markovic, E. Alon and V. Stojanovic, "Demonstration of integrated micro-electro-mechanical relay circuits for VLSI applications," *IEEE J. Solid-State Circuits*, vol. 46, pp: 308-320, Jan. 2011.

[10] C. Ayala, D. Grogg, A. Bazigos, S. Bleiker, M. Fernandez-Bolaños, F. Niklaus, and C. Hagleitner, "Nanoelectromechanical digital logic circuits using curved cantilever switches with amorphous-carbon-coated contacts," *Solid-State Electronics*, 2015 (in press).

[11] R. Venkatasubramanian, S. Manohar, and P. Balsara, "NEM relay based sequential logic circuits for low power design," *IEEE Trans. Nanotechnology*, vol. 12, no. 3, pp. 386–398, May 2013.

[12] A. Bazigos, C. Ayala, S. Rana, D. Grogg, M. Fernandez-Bolanos, C. Hagleitner, T. Qin, D. Pamunuwa and A. Ionescu, "Analytical Compact Model in Verilog-A for Electrostatically Actuated Ohmic Switches", *IEEE Trans. on Electron Devices*, vol. 61, pp. 2186-2194, Jun. 2014.

[13] S. Dutta and V. Stojanovic. "Floating-point unit design with nano-electro-mechanical (NEM) relays," *IEEE/ACM Int. Symp. Nanoscale Architectures*, Paris, 2014, pp. 145-150.

[14] S. Rana, T. Qin, A. Bazigos, D. Grogg, M. Despont, C. Ayala, C. Hagleitner, A. M. Ionescu, R. Canegallo and D. Pamunuwa, "Energy and Latency Optimization in NEM Relay-Based Digital Circuits," *IEEE Trans. Circuits and Systems I*, vol. 61, pp.2348-59, Aug. 2014.

[15] G. Ayers at el. (2012), eXtensible Utah Multicore (XUM) Project[online]. Available: https://github.com/grantea/mips32r1

[16] S. Rana, T. Qin, D. Grogg, M. Despont, Y. Pu, C. Hagleitner, and D. Pamunuwa, "Modelling NEM relays for digital circuit applications," *IEEE Int. Symp. Circuits and Systems*, Beijing, 2013, pp. 805-808.

[17] L. A. Bumm, J. J. Arnold, T. D. Dunbar, D. L. Allara and P. S. Weiss, "Electron Transfer through Organic Molecules," *J. Phys. Chem. B*, vol. 103, pp 8122–8127, 1999.

[18] B. Ackalloor and D. Gaitonde, "An Overview of Library Characterization in Semi-Custom Design," *Proc. IEEE Custom Integrated Circuits Conf.*, Santa Clara, CA, 1998, pp. 305-312.

[19] R. Pollack. (2012). 4004 CPU and MCS-4 family chips Project [online]. Available: http://opencores.org/project,mcs-4.

[20] C. Zhang, H. Yu, Wei. Zhang, "A Nano-Electro-Mechanical-Switch based Thermal Management for 3D Integrated Many-core Memory-Processor System", *IEEE Trans. Nanotechnology*, vol.11, no.3, pp.588-600, May 2011

[21] H. Fariborzi, F. Chen, V. Stojanovic, R. Nathaneal, J. Jeon, Tsu-Jae.K.Liu, "Design and Demonstration of Micro-Electro-Mechanical Relay Multipliers", *Proc. IEEE Asian Solid State Circuits Conf.*, Jeju, South Koera, 2011, pp. 117-120.